Regional Precipitation Quantile Values for the Continental United States Computed from L-Moments

NATHANIEL B. GUTTMAN

National Climatic Data Center, Asheville, North Carolina

J. R. M. HOSKING AND JAMES R. WALLIS

IBM T. J. Watson Research Center, Yorktown Heights, New York

(Manuscript received 8 September 1992, in final form 16 April 1993)

ABSTRACT

Precipitation quantile values have been computed for 9 probabilities, 8 durations, 12 starting months, and 111 regions across the United States. L-moment methodology has been used for the calculations. Discussed are the rationale for selecting the Pearson type III (gamma) and Wakeby distributions, and the confidence that can be placed in the quantile values. Results show that distribution functions become more asymmetrical as the duration decreases, indicating that the median may be a better measure of central tendency than the mean. Portraying the quantile values as a percentage of the median value leads to smooth spatial fields.

Computation of quantile values was the first known large-scale application of L-moment methodology. In spite of the complexity of the techniques and the extensive use of personnel and computer resources, the results justify the procedures in terms of preparing easy to use probability statements that reflect underlying physical processes.

1. Introduction

A national drought atlas is being prepared jointly by the U.S. Army Corps of Engineers, National Climatic Data Center, U.S. Geologic Survey, and IBM Thomas J. Watson Research Center (Willeke et al. 1991; Guttman et al. 1991; Guttman 1993). In order to accommodate a variety of needs, it was decided that part of the portrayal of drought should be in terms of precipitation probabilities. For durations of 1, 2, 3, 6, 12, 24, 36, and 60 months, beginning in each calendar month January through December, regional quantile values for probabilities of 0.02, 0.05, 0.10, 0.20, 0.50, 0.80, 0.90, 0.95, and 0.98 have been calculated using L-moment methodology. A complete description of L-moments is given by Hosking (1990).

The statistical methods involve regional analysis. Using "index flood" procedures (Hosking and Wallis 1991; Kite 1988), it is assumed that throughout a region the distribution of precipitation is the same at all sites except for a scale factor that may vary from site to site. A weighted-average regional dimensionless quantile function or growth curve is computed, and the site-specific quantile estimates are obtained by multiplying the site mean (scale factor) by the regional quantile function.

Corresponding author address: Dr. Nathaniel B. Guttman, National Climatic Data Center, Federal Building, Asheville, NC 28801.

This paper concerns the use of L-moments in calculating the regional quantile values. Guttman (1993) described L-moments and their use in defining 104 regions within the 48 contiguous United States. The number of regions in the National Drought Atlas for which quantiles were estimated was increased to 111 for reasons that will be given later. The regions contain between 1 and 48 sites; half of the sites portrayed in the atlas are in regions containing at least 13 sites. The median region size is 8 sites with 645 station years of data.

2. Distribution fitting

If a region is homogeneous in the sense that the data for each site within the region represent a random realization of the same underlying physical process, then the regional weighted-average L-moments can be used to fit a probability distribution (Hosking 1990; Hosking and Wallis 1990; Wallis 1989). The weighting is proportional to the number of data values at each site.

Regional average L-moments were computed and used to fit the three-parameter generalized extreme value, Pearson type III (gamma), generalized logistic (as defined by Hosking 1990, Table 1), and lognormal distributions. Two-parameter distributions were not considered. because the regions are typically large enough so that the third parameter can be estimated with sufficient accuracy (as indicated by the root-mean-square error results discussed in section 4). A measure constructed by Hosking and Wallis (1991) was used to evaluate the goodness of fit. This measure is based on the difference between L-kurtosis of the fitted distribution and the regional average L-kurtosis of the sample data. Assessment of goodness of fit is based on L-kurtosis, the fourth L-moment, because the first three L-moments are used to estimate the three parameters of the distribution.

Counts were made by duration, region, and starting month of the number of times a distribution was acceptable. Table 1 summarizes the counts over all regions by duration and starting month. The gamma was found to be acceptable most often for precipitation totals over all durations. The lognormal and generalized extreme value distributions were acceptable almost as often as the gamma for durations longer than six months.

Of the four three-parameter distributions, the generalized logistic was acceptable least often. This distribution could only be used in less than 20% of the regions for durations of a year or less. For longer durations, the acceptance rate increased as the time period increased; the rate was near 30% for two-year precipitation totals and about 60% for five-year totals.

As seen in Table 1, in many regions more than one distribution passed the goodness-of-fit test. This means that the amount of data in the region was not sufficient to enable discrimination between all the distributions. This is not surprising since some of the distributions closely resemble each other over certain ranges of skewness. At the low skewness values typical of 12-

TABLE 1. Distributions accepted by the goodness-of-fit test, by duration and starting month over 104 regions. Distributions tested are generalized logistic (Gen Log), generalized extreme value (Gen Ev), lognormal (Lognorm), and gamma. The ">1" rows give the number of regions for which more than one distribution was accepted.

Duration		Starting month											
(months)	Distribution	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	Gen Log	47	26	29	26	27	23	29	26	41	42	26	38
	Gen Ev	34	37	42	40	44	44	26	25	29	37	34	32
	Lognormal	49	44	45	49	49	49	37	36	50	55	41	42
	Gamma	55	53	55	57	64	59	53	47	68	70	58	46
	>1	60	48	49	52	53	52	43	41	55	64	50	48
2	Gen Log	24	27	27	26	24	24	15	20	37	26	28	34
	Gen Ev	49	49	53	50	61	47	46	47	63	52	57	50
	Lognorm	53	52	54	54	57	52	51	58	71	58	59	52
	Gamma	53	58	56	57	61	51	51	67	66	66	65	55
	>1	56	56	55	53	60	50	51	60	73	60	63	57
3	Gen Log	19	26	27	25	25	14	14	27	20	29	27	20
	Gen Ev	52	58	63	50	60	45	53	68	53	52	58	54
	Lognorm	55	56	62	59	64	47	52	72	62	54	60	54
	Gamma	57	53	61	55	58	52	60	73	66	57	60	56
	>1	56	59	65	58	66	48	56	74	63	56	63	55
6	Gen Log	20	22	20	18	18	15	12	19	19	15	22	18
	Gen Ev	57	56	52	61	64	68	64	62	61	48	48	55
	Lognorm	60	60	57	62	71	71	68	66	60	57	53	60
	Gamma	63	61	57	64	69	69	66	66	63	55	54	57
	>1	64	62	58	64	70	70	67	66	61	56	52	60
12	Gen Log	19	16	19	14	21	16	12	19	20	14	21	22
	Gen Ev	78	68	63	60	55	65	65	66	66	65	73	71
	Lognorm	84	76	71	62	67	68	62	65	72	64	76	81
	Gamma	86	76	74	64	65	69	63	65	72	71	75	77
	>1	85	78	73	63	66	69	65	68	73	70	79	81
24	Gen Log	31	31	30	25	32	30	29	32	31	27	27	34
	Gen Ev	78	79	76	81	80	83	75	78	76	73	77	77
	Lognorm	80	79	80	82	83	83	76	81	77	75	77	79
	Gamma	77	80	79	83	83	82	74	82	77	76	77	77
	>1	80	81	80	83	83	83	75	82	78	77	78	79
36	Gen Log	45	41	35	37	43	50	45	42	43	43	43	38
	Gen Ev	71	72	69	75	79	80	73	74	78	82	78	78
	Lognorm	74	75	67	69	77	78	74	76	79	77	76	80
	Gamma	75	76	69	71	79	79	74	76	78	75	76	80
	>1	75	76	69	71	79	80	75	77	79	77	76	80
60	Gen Log	60	60	67	62	64	59	61	61	54	61	57	54
	Gen Ev	73	79	74	74	77	76	70	72	71	66	67	72
	Lognorm	74	79	79	80	79	82	72	74	74	71	71	72
	Gamma	74	79	78	81	78	83	72	74	73	71	69	72
	>1	74	79	79	81	80	83	72	75	75	71	71	72

month precipitation, for example, the lognormal and gamma distributions are very similar; they both reduce to the normal distribution when the skewness is zero. When more than one distribution is accepted by the goodness-of-fit test, the estimated quantiles may be expected to be very similar except in the extreme tails of the distributions. In our study we were concerned not with extreme tail quantiles, but only with quantiles in the range 0.02 to 0.98, and the differences between estimated quantiles were generally small compared with the root-mean-square errors of the quantile estimates (these root-mean-square errors are discussed in section 4). We therefore deemed it adequate to use any of the distributions that passed the goodness-of-fit test.

Because of user friendliness considerations about the atlas, it was decided to compute quantile values, if possible, from only one distribution function for all regions and durations. Based on the counts of acceptable fits, the gamma was chosen. However, the gamma was not acceptable for all time periods and regions.

Two conditions preclude the use of the gamma. First, the goodness-of-fit measure finds it unacceptable. Second, the region is heterogeneous. Homogeneity depends on the time period for which quantiles are desired. Only three of the 104 regions were heterogeneous in terms of calendar-year precipitation amounts, but as expected, the number of heterogeneous regions [as determined from the procedures described by Guttman (1993)] increased as the durations and starting month of the durations changed. Since the regions were constrained to be fixed for all time periods, the question arose as to how quantiles should be computed for heterogeneous regions.

For this second condition there is no reason to assume that a single distribution will give a good fit to every site's data within the heterogeneous region. The Wakeby five-parameter distribution was chosen as the single fitted distribution for a heterogeneous region. The Wakeby was also chosen as the distribution for a region where the gamma was unacceptable. The generalized extreme value, generalized logistic, and lognormal were not chosen because they were rarely acceptable when the gamma was unacceptable.

The Wakeby distribution is defined as a probability distribution whose quantile function or inverse cumulative distribution function, $x(F)$, is

$$x(F) = \xi + (\alpha/\beta)[1 - (1 - F)^{\beta}] - (\gamma/\delta)[1 - (1 - F)^{-\delta}]$$

where F is the cumulative distribution function, and ξ, α, β, γ, δ are real valued parameters. It was chosen as a default distribution because it is robust to misspecification of the underlying distribution function for a region (Kotz et al. 1988; Hosking and Wallis 1991), and it can attain a wide range of distributional shapes with fixed lower bounds that mimic many skew distributions, making it particularly useful for hydrologic studies (Hosking 1986).

Once a distribution was chosen, quantile values were calculated by fitting the distribution from the regional average L-moments. In dry areas for the shorter durations some of the values were negative. Since precipitation amounts are calculated by multiplying a site or regional mean precipitation amount by a quantile value, negative values violate the physical lower bound of zero precipitation totals. The problem was solved by applying a mixed model

$$F(x) = p + (1 - p)G(x)$$

where F is the cumulative distribution function (cdf) of precipitation amounts, p is the probability that the precipitation amount is zero as estimated by the proportion of zero values in the data for the region, and G is the cdf of the distribution of nonzero precipitation amounts as estimated from the regional average L-moments of the nonzero data values. Note that for distribution fitting, the L-moments were computed from only the nonzero data; L-moments computed from both the nonzero and zero data were used for defining regions.

As stated previously, the distribution G was initially chosen to be gamma in homogeneous regions for which the gamma distribution was accepted by the goodness-of-fit criterion, and Wakeby otherwise. However, G was constrained to have a lower bound of zero when this was necessary to obtain nonnegative quantiles for all the probabilities of interest (the lowest of these is 0.02). When constrained estimation was necessary, the Wakeby with fixed lower bound $\xi = 0$ was fit. A gamma distribution with zero lower bound was not used because it has only two free parameters, and it rarely gave a good fit to the data. The lognormal and generalized extreme value distributions were not reconsidered as a suitable model because the Wakeby can mimic the shapes of these distributions. The algorithm that was used to choose a distribution is shown in Fig. 1.

The Wakeby distributions were fit using Hosking's (1991) implementation of the algorithm of Landwehr et al. (1979). In this implementation, when the full five-parameter Wakeby distribution cannot be fit, successive attempts are made to fit the Wakeby with lower bound $\xi = 0$, with $\alpha = \beta = 0$ or $\gamma = \delta = 0$ (also known as the generalized Pareto), and with $\xi = \alpha = \beta = 0$ or $\xi = \gamma = \delta = 0$ (also known as the Pareto distribution) until a successful fit is achieved. Goodness-of-fit tests were not applied to the Wakeby because the distribution has the ability to attain the wide variety of skewed shapes that the cumulative precipitation distributions exhibit.

Quantile values for the 104 regions defined by Guttman (1993) for a given duration, starting month, and quantile were mapped and subjectively examined for geographical consistency. For several durations and starting months, there was a large gradient across four areally extensive regions in the plains and Southeast. In order to better define the gradients, these four regions were subdivided using the procedures described by

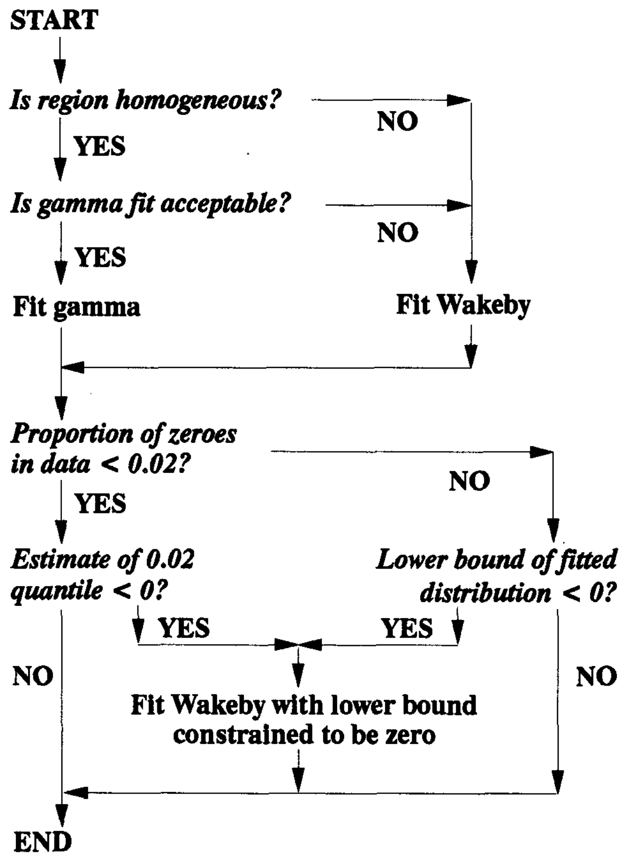

FIG. 1. Algorithm for choosing a distribution.

Guttman (1993). The original 104 regions were thus increased to 111 regions. Distributions were fit to the new regions, quantile values were recomputed, and new maps were prepared. For a duration starting in January, Table 2 shows the number of regions by duration that are homogeneous, as well as the number that are fit by gamma and Wakeby distributions.

3. Quantile estimates

Values have been computed for 9 quantiles (0.02 to 0.98) for 111 regions for 8 durations (1 to 60 months) for 12 time-period starting months (January to December). Computer files of the site information, input data, L moments, quantile values, precipitation amounts, and distribution parameter estimates are available from the National Climatic Data Center. It is anticipated that the information will also be available on CD-ROM.

Since the purpose of the computations was to characterize drought, the discussion of results is restricted to patterns of the 0.02 to 0.50 quantiles. For durations of a year or longer, the spatial patterns of a quantile value expressed as a percentage of either the mean or median exhibit very little change as the starting month of the duration series slides from January through December. The 12- through 60-month precipitation totals, starting in January, can therefore be considered representative of the totals for other starting months.

Results are depicted on "starburst" maps (Figs. 2–16). Lines radiate from the centroid of a region to the location of sites that make up the region. Regional statistics are plotted at the centroid of a region. Heterogeneous regions are indicated by parentheses around the regional statistic. The maps show both the geographic extent of and the number of sites within a region. The depiction clearly indicates areas with sparse data coverage as well as areas not covered by a region. Explicit regional boundaries are not defined because interpolation between regions may require a more detailed local knowledge of precipitation climates.

Figure 2 shows the median (0.50 quantile) expressed as a percentage of the mean for the 12-month precipitation totals. This ratio not only provides an indication of the skewness of the precipitation distributions, but also is useful in converting the quantile scale factor from the mean to the median to obtain a precipitation amount from the quantile values. Except for the arid regions of the Southwest, the median is generally within 2% of the mean annual precipitation. The largest difference between the median and mean, 13%, is in the southern California desert. In most of the other arid regions of the Southwest the difference is about 3% to 5%. As the duration increases, the same pattern is evident, but the differences decrease; in the southern California desert the difference for the 60-month duration is only 5%.

As the duration decreases, the median becomes increasingly smaller than the mean indicating a progressive asymmetry of the distributional shape. This result is especially true in arid regions, but is evident throughout the country. Figures 3 and 4 show examples for the January one-month precipitation and the January through March three-month totals. The degree of asymmetry also follows the annual march of precipitation in a region. As shown in Figs. 4–7, for example, the drier seasons show more of a departure from symmetry than the wetter seasons. This temporal pattern is magnified as the duration decreases.

Comparisons of the mean and median indicate that the median may be a better representation of the central

TABLE 2. Number of gamma and Wakeby distribution fits for durations beginning in January.

Duration (months)	Homogeneous (gamma)	Homogeneous (Wakeby)	Heterogeneous (Wakeby)
1	46	46	19
2	52	36	23
3	60	26	25
6	63	28	20
12	92	16	3
24	83	21	7
36	78	25	8
60	78	23	10

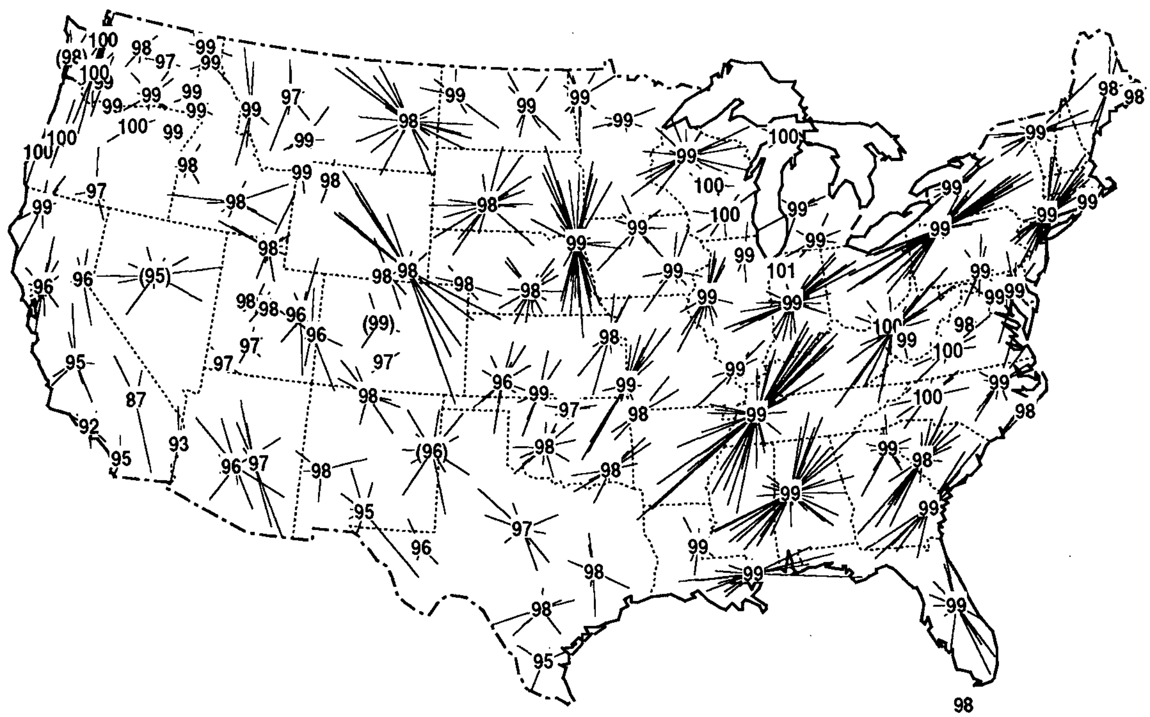

FIG. 2. Median as a percentage of the mean for 12-month precipitation January through December. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

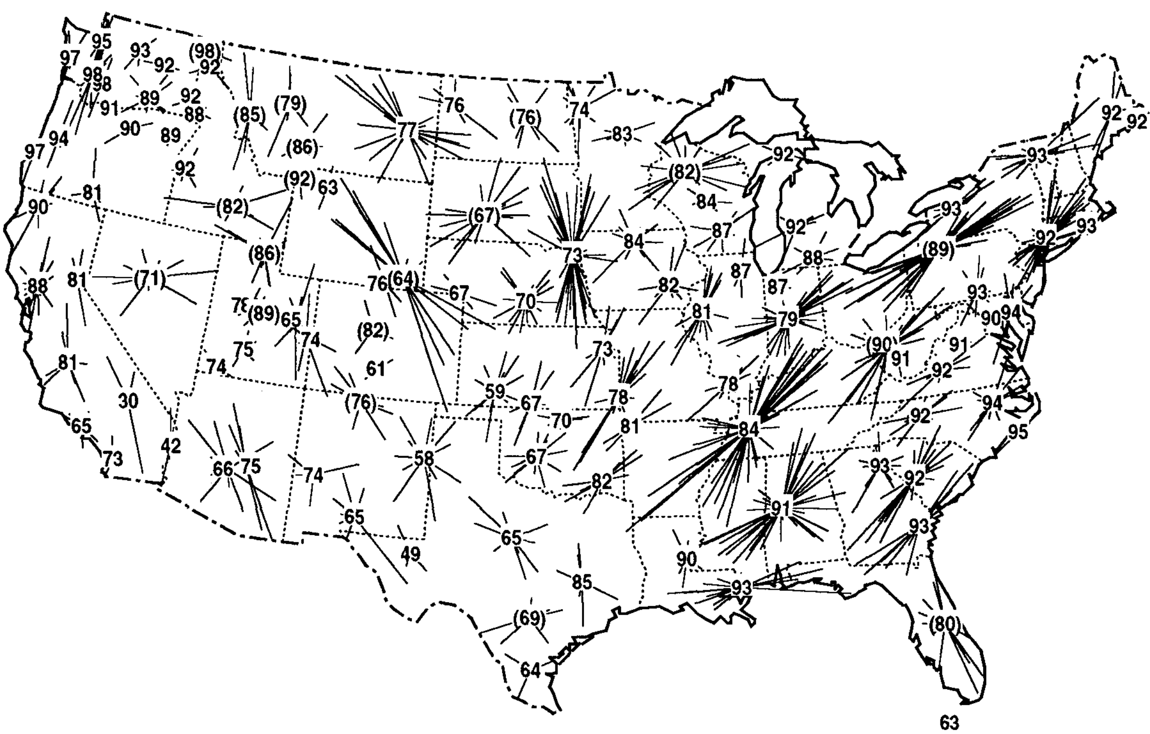

FIG. 3. Median as a percentage of the mean for January total precipitation. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

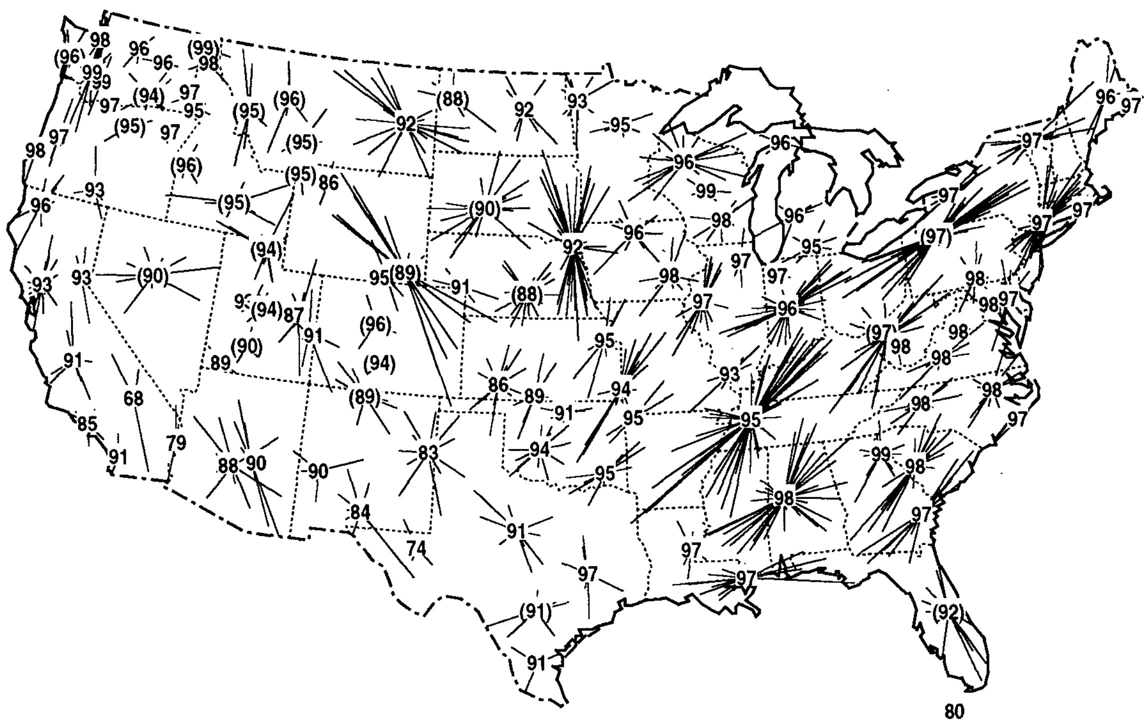

FIG. 4. Median as a percentage of the mean for three-month total precipitation January through March. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

tendency of precipitation than the mean. Historically, climatic normals have been used extensively to compare current to expected conditions. Normal precipitation is based on mean values, and the popular interpretation is that departures from a mean are symmetric (Guttman 1989). Because of the asymmetry of the precipitation distributions, the use of the median as an expected central tendency would reduce the effects of the popular misconceptions about normal precipitation.

Temporal and spatial patterns in the quantile values expressed as a percentage of the 0.50 quantile value (median) become evident as the duration decreases from 12 months to 1 month. Figures 8 through 15 depict the patterns for the 0.10 quantile for the three-month totals for the meteorological seasons and the one-month totals for the midseason months. These maps are representative of the patterns that evolve as the duration and its starting month change. Although the magnitude of the percentages changes as the quantile changes, quantiles other than 0.10 show the same patterns as depicted in the figures. Results are presented in terms of the ratio of a quantile to the median to simplify pattern recognition; presentations in terms of precipitation amounts would contain scale effects that are eliminated by the ratios.

The maps indicate heterogeneous regions by parentheses surrounding a percentage. It is seen from the figures and also in Table 2 that for annual precipitation, 108 out of the 111 regions are homogeneous. For the shorter durations about 70% to 90% of the regions are homogeneous. The larger degree of heterogeneity for the shorter durations results from the fact that regions were determined from the annual pattern of precipitation rather than from duration- and time of year–dependent precipitation characteristics. Despite the heterogeneity, and except for some mountainous areas in the West, the percentages form smooth spatial fields. This result suggests that because the transitions among neighboring regions are smooth, as would be expected from physical considerations, the L-moment procedures do indeed describe the physical processes that operate in each region.

4. Accuracy of estimates

Quantile values were assessed by their bias and root-mean-square error (RMSE). These quantities cannot be calculated analytically because the regional L-moment quantile estimation procedure is too complicated. Instead, a Monte Carlo simulation procedure was used. Simulated data were generated for a region with the same number of sites and the same record lengths as the actual region, and were drawn from the distribution that was fit to the actual regional data. Quantile estimates were calculated for the sites in this simulated region. The simulation was repeated 500 times. The 500 sets of errors in the simulated quantile estimates

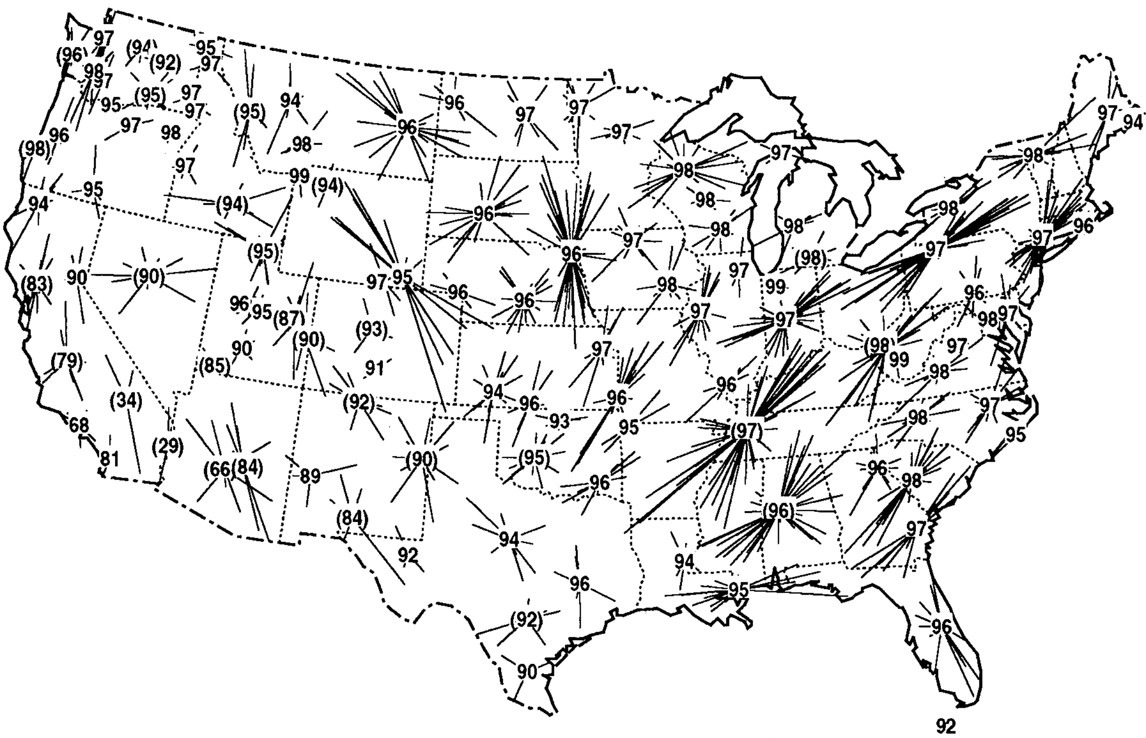

FIG. 5. Median as a percentage of the mean for three-month total precipitation April through June. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

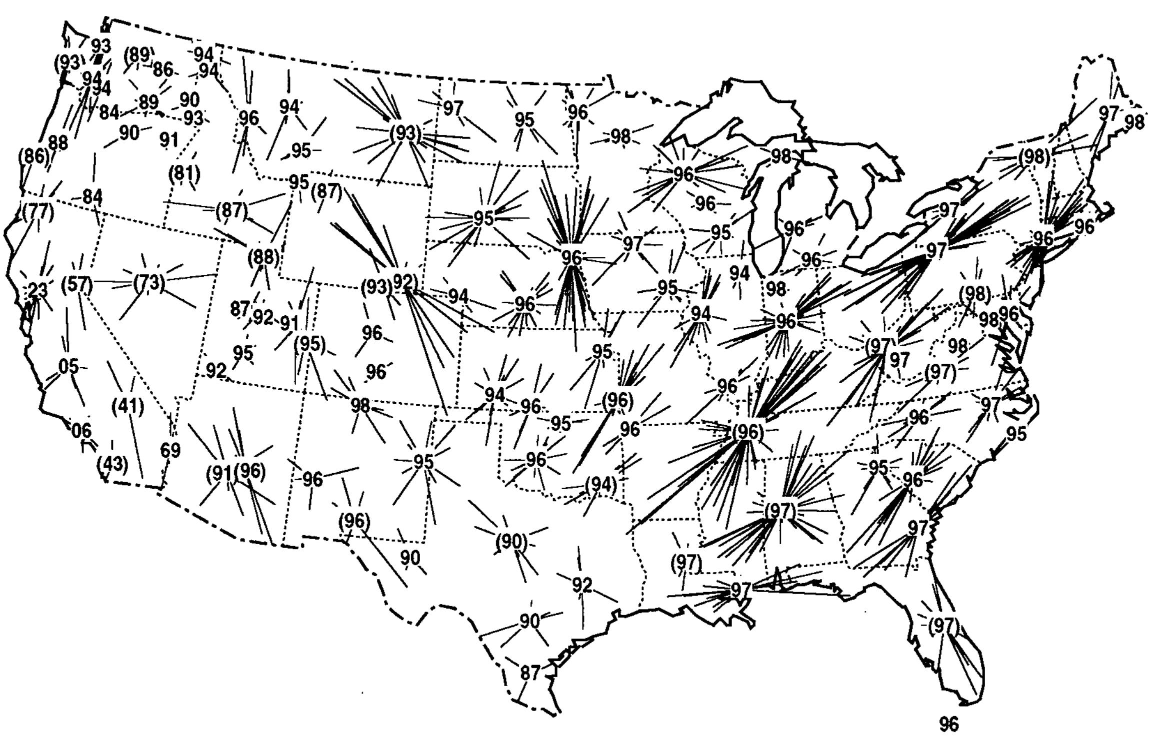

FIG. 6. Median as a percentage of the mean for three-month total precipitation July through September. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

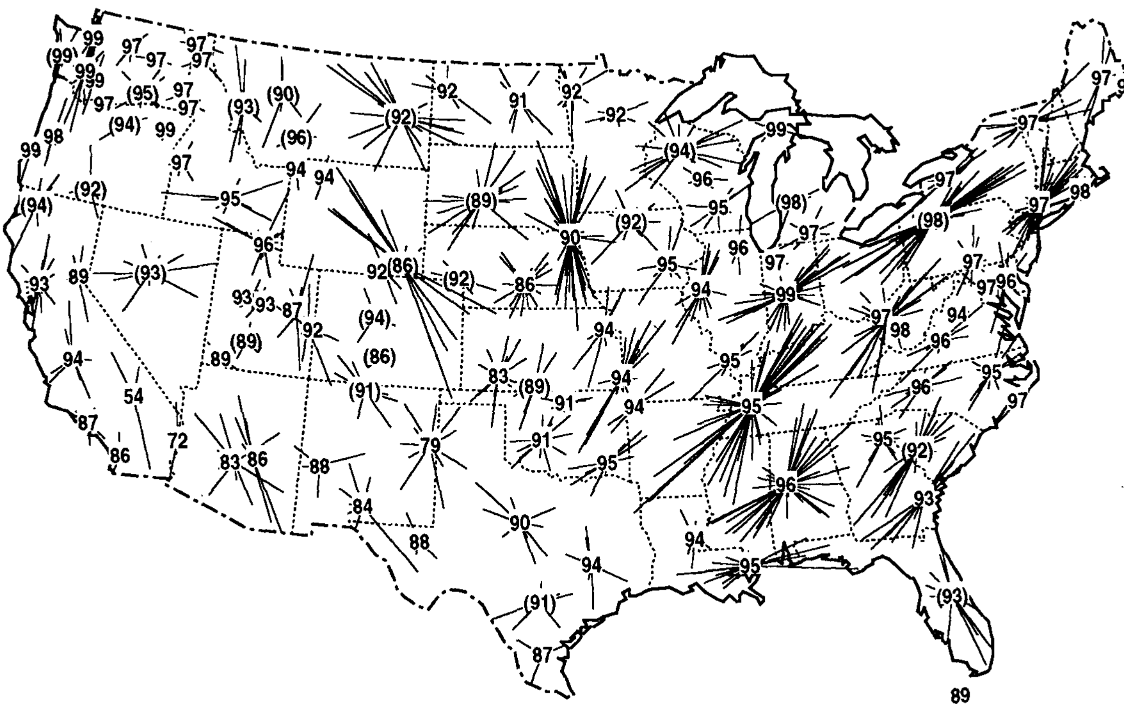

FIG. 7. Median as a percentage of the mean for three-month total precipitation October through December. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

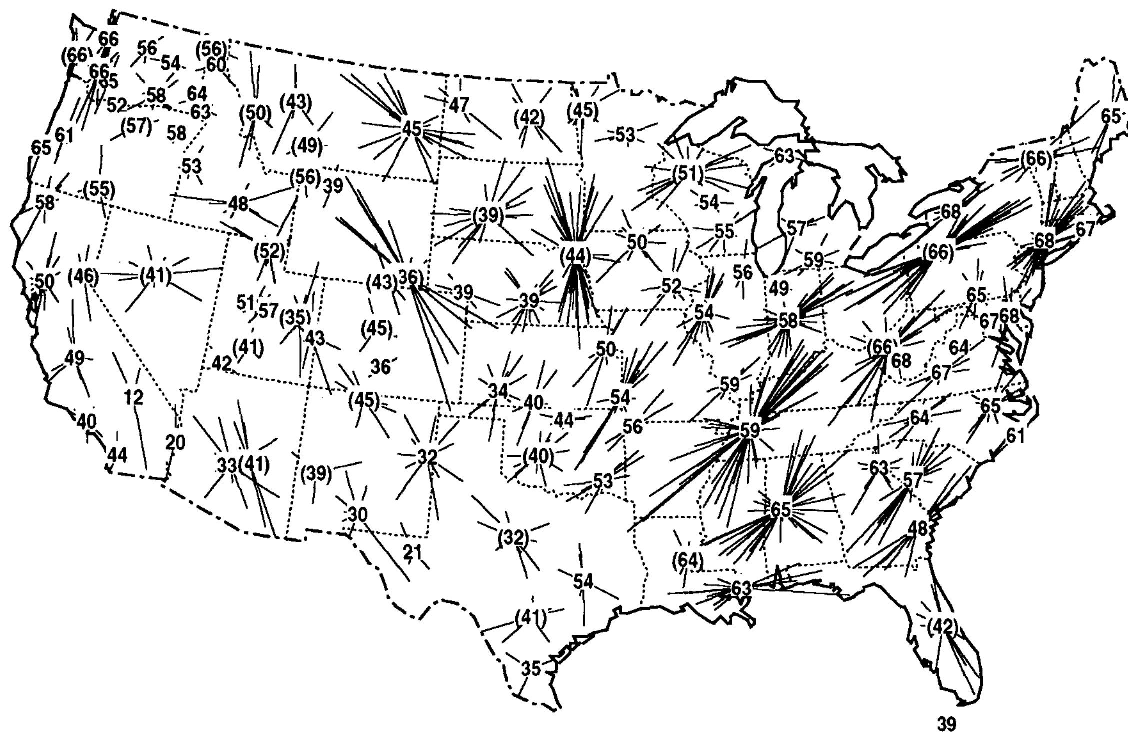

FIG. 8. The 0.10 quantile value expressed as a percentage of the 0.50 quantile value (median) for three-month total precipitation December through February. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

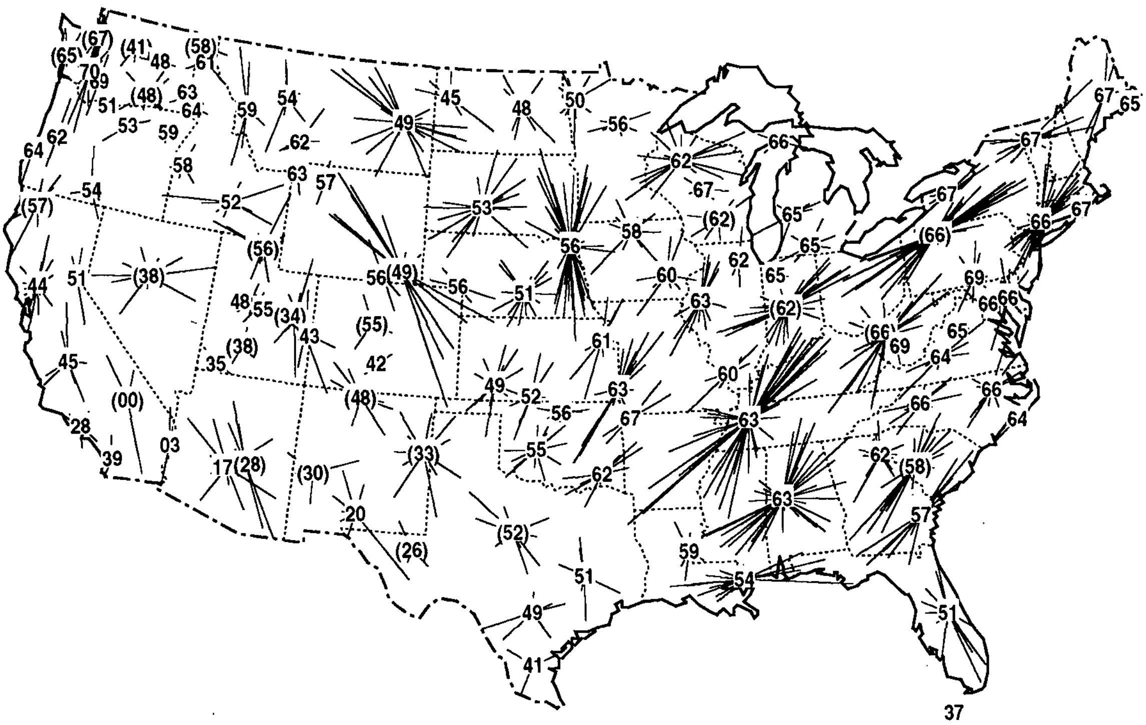

FIG. 9. The 0.10 quantile value expressed as a percentage of the 0.50 quantile value (median) for three-month total precipitation March through May. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

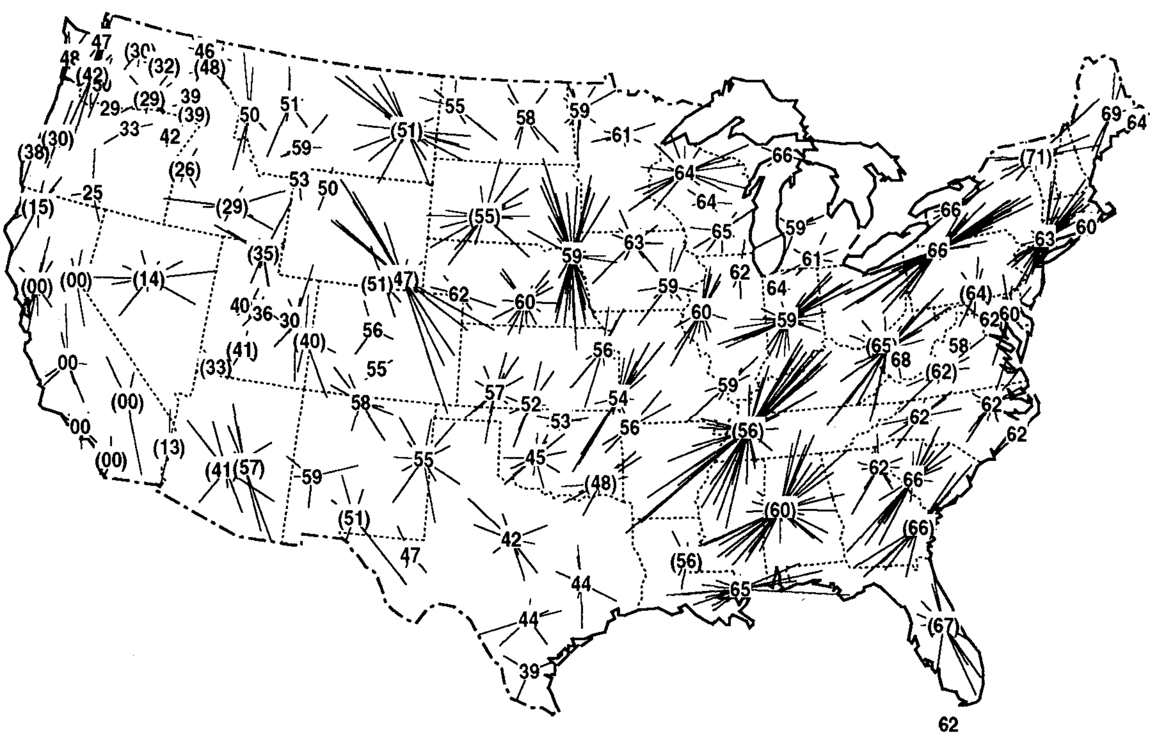

FIG. 10. The 0.10 quantile value expressed as a percentage of the 0.50 quantile value (median) for three-month total precipitation June through August. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

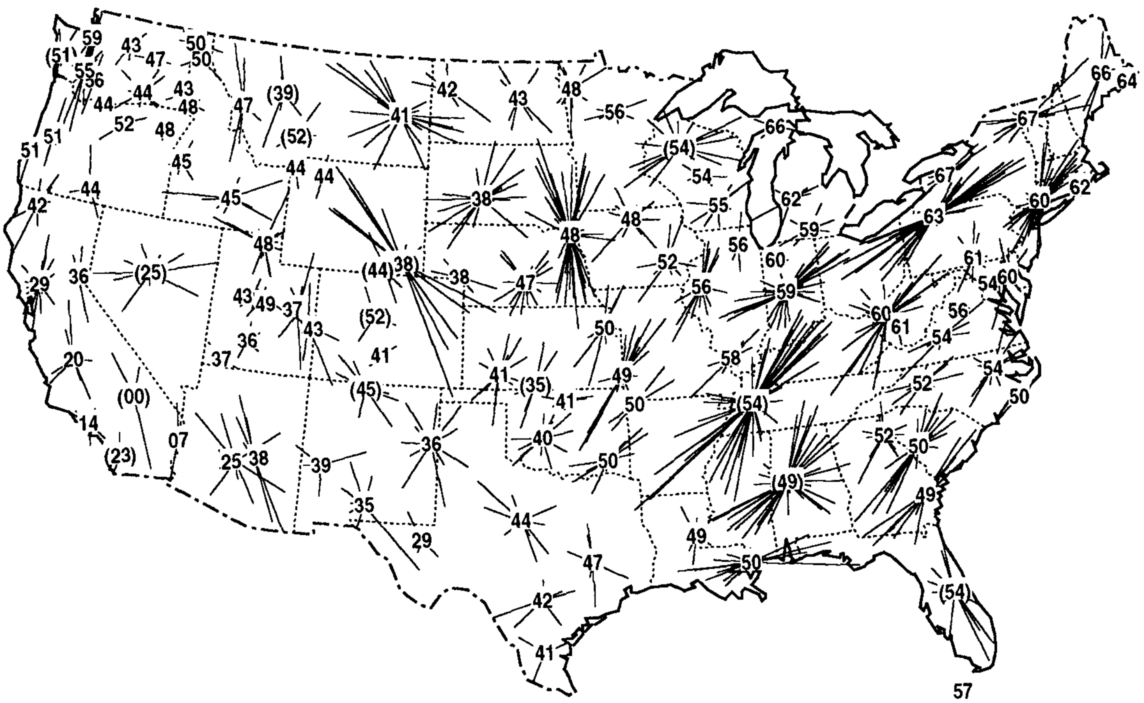

FIG. 11. The 0.10 quantile value expressed as a percentage of the 0.50 quantile value (median) for three-month total precipitation September through November. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

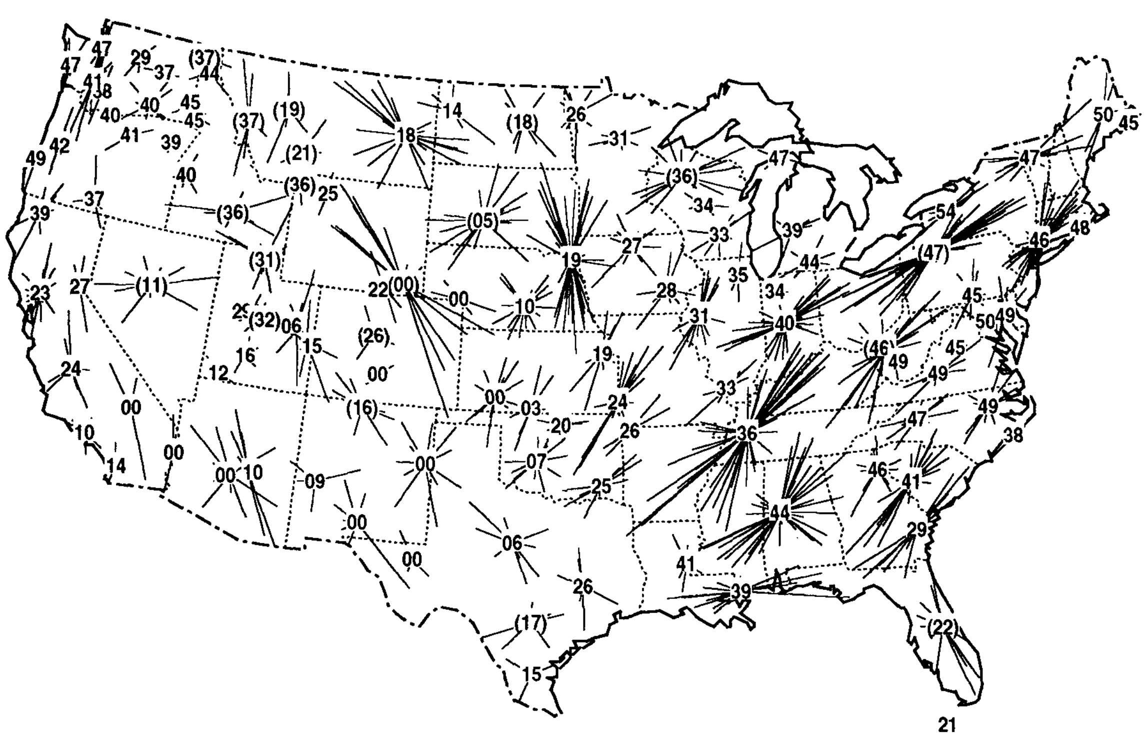

FIG. 12. The 0.10 quantile value expressed as a percentage of the 0.50 quantile value (median) for January total precipitation. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

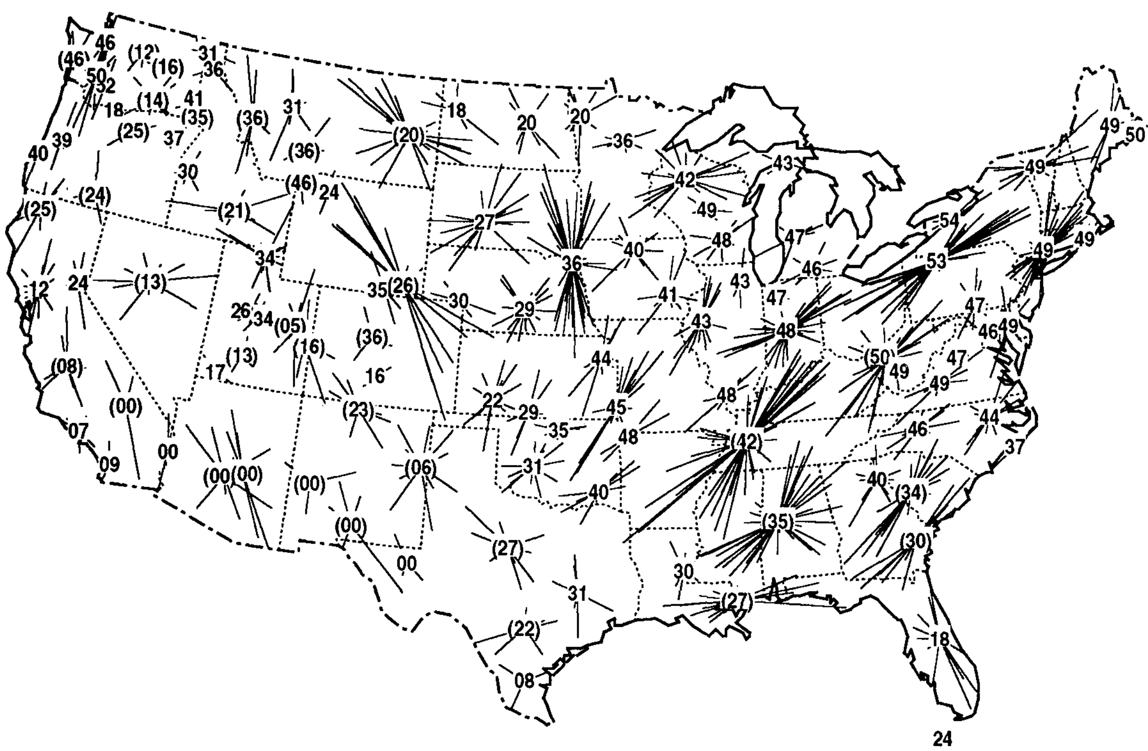

FIG. 13. The 0.10 quantile value expressed as a percentage of the 0.50 quantile value (median) for April total precipitation. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

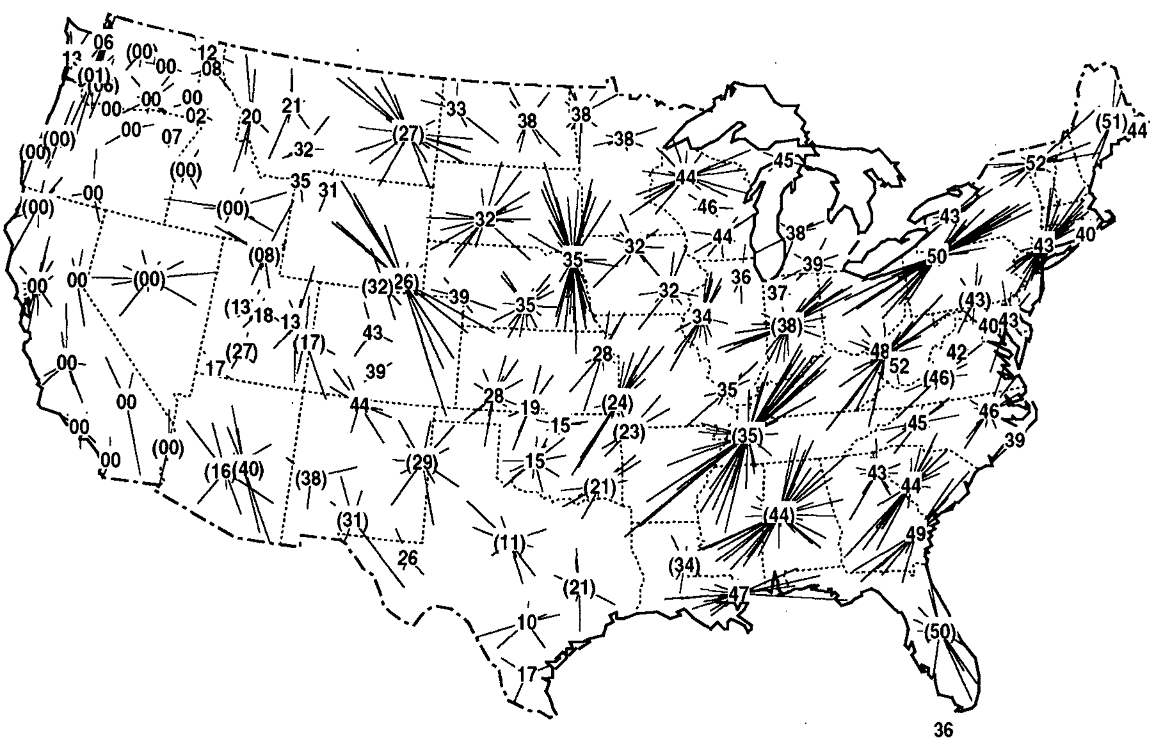

FIG. 14. The 0.10 quantile value expressed as a percentage of the 0.50 quantile value (median) for July total precipitation. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

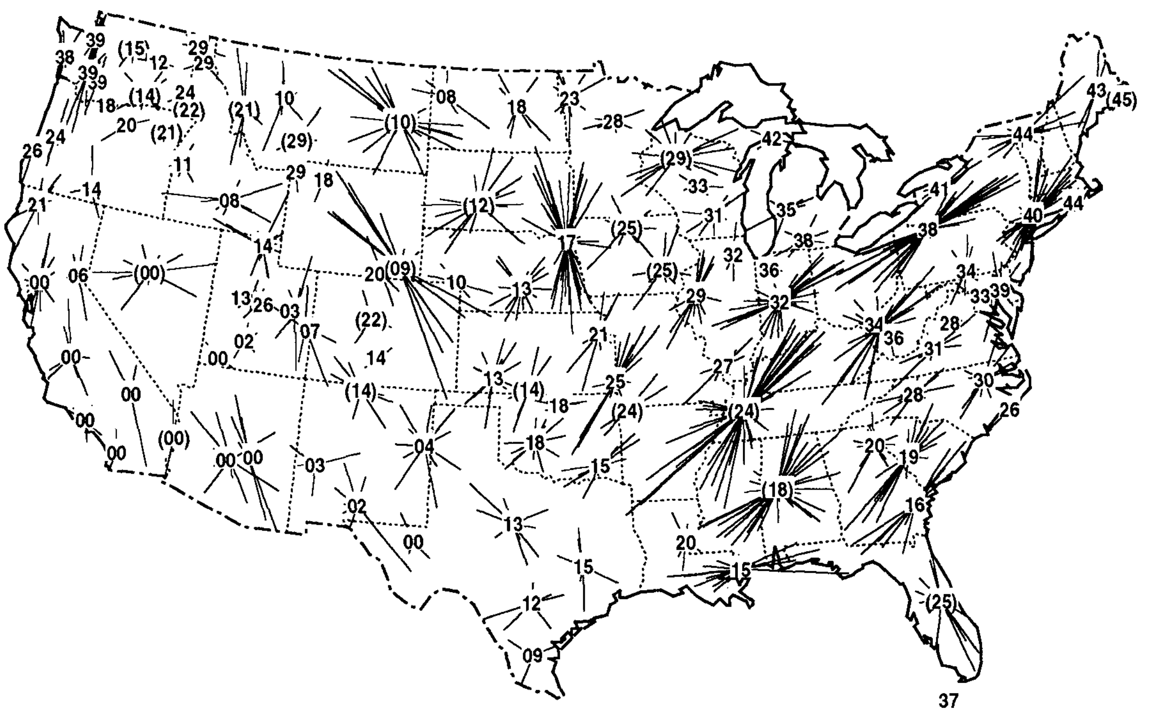

FIG. 15. The 0.10 quantile value expressed as a percentage of the 0.50 quantile value (median) for October total precipitation. The number in the centroid of a region indicates the percentage; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the percentage.

were accumulated and averaged to yield approximations to the bias and RMSE of the quantile estimates calculated from the actual data.

For all durations, for quantiles between 0.02 and 0.50, the bias is negligible, and the RMSE is less than 0.10. For the shorter durations, for quantiles greater than 0.50, the bias and RMSE are only slightly larger. The bias and standard error generally decrease as the duration increases.

The main exceptions are those instances in which a fitted Wakeby distribution had three zero parameters (i.e., a Pareto distribution). This occurred when fitting a few of the shorter-duration samples. In these cases at the higher quantiles, especially 0.98, quantile estimates have high RMSE and substantial negative bias. Although one can have little confidence in these quantile estimates, the values are conservative for purposes of drought planning.

The bias and RMSE were computed with the assumption that there is no intersite correlation within a region. Since the correlation between sites can typically be between 0.4 and 0.8, simulations were run for six representative regions to compare the bias and RMSE for correlated sites with uncorrelated sites. Five hundred simulations were run for homogeneous regions for annual precipitation and for January precipitation (duration of one month). The methods of generating the intersite correlations and then the simulated data are described by Hosking and Wallis (1988, p. 590).

Table 3 summarizes the regions. Tables 4 and 5 give the bias and RMSE of regional quantile estimates 1) with no correlation between sites and 2) with the average intersite correlation being the same as for the real-world region. The amount of correlation present in the data has a negligible effect on the bias of quantile estimates but inflates the RMSE by a factor of 2 to 3. This should not be a problem in practice, since most estimates will still have an RMSE of 0.03 or less.

As a further example, RMSE values were computed, taking intersite correlation into account, for all of the regions for quantiles of precipitation totals for a duration of the three months December through February. The RMSE's for the ratio of the 0.10 quantile to the median are shown in Fig. 16; these values may be regarded as the standard errors of the estimates mapped in Fig. 8. The RMSEs are certainly small enough to enable the quantile estimates to be used with confidence.

Accuracy measures are portrayed solely as the bias and rmse that were determined from the simulation. Computer files of the measures are available from the National Climatic Data Center. Confidence intervals have not been computed. Following usual practice, intervals could be constructed by adding and subtracting the product of the rmse of the estimated quantile value and the appropriate percentage point from the standard normal distribution to the quantile estimate. This construction assumes that quantile estimates are normally distributed. The validity of this assumption is, however, dubious for extreme quantiles, for arid areas, and for quantiles close

TABLE 3. Location, number of sites, probability distribution, and average intersite regional correlation for simulations.

Region number	Location	Number of sites	Annual Distribution	Annual Average correlation	January Distribution	January Average correlation
10	Northeast coast	36	Gamma	0.63	Wakeby	0.79
106	Alabama, Georgia	15	Gamma	0.69	Gamma	0.76
49	Minnesota, Iowa	14	Gamma	0.59	Gamma	0.67
27	Central Texas	10	Gamma	0.58	Wakeby	0.71
60	Pacific NW mountains	3	Wakeby	0.80	Wakeby	0.83
82	California (San Joaquin Valley)	9	Gamma	0.75	Gamma	0.76

to zero. It may also be questionable for other quantiles and other areas. Unless the assumption of normality of quantile estimates is verified, the usual practice of constructing confidence intervals is strongly discouraged.

5. Evaluation

Regional weighted average L-moments have been used to estimate quantile values of precipitation. Except for a few isolated cases, confidence in the estimates, as

TABLE 4. Simulation results for annual precipitation.

Probability	0.02	0.05	0.10	0.50	0.80
Region 10 quantiles	0.697	0.749	0.798	0.991	1.134
Region 10 uncorrelated					
bias	−0.005	−0.005	−0.004	−0.001	0.003
RMSE	0.007	0.006	0.005	0.002	0.003
Region 10 correlated					
bias	−0.005	−0.004	−0.004	−0.001	0.003
RMSE	0.016	0.013	0.010	0.004	0.007
Region 106 quantiles	0.676	0.729	0.780	0.986	1.146
Region 106 uncorrelated					
bias	−0.005	−0.005	−0.004	−0.001	0.003
RMSE	0.009	0.007	0.006	0.003	0.004
Region 106 correlated					
bias	−0.005	−0.005	−0.005	−0.001	0.003
RMSE	0.020	0.015	0.012	0.005	0.008
Region 49 quantiles	0.613	0.682	0.746	0.990	1.168
Region 49 uncorrelated					
bias	−0.002	−0.003	−0.003	−0.002	0.002
RMSE	0.010	0.008	0.006	0.003	0.004
Region 49 correlated					
bias	−0.001	−0.002	−0.003	−0.002	0.002
RMSE	0.021	0.016	0.012	0.005	0.008
Region 27 quantiles	0.504	0.582	0.658	0.974	1.228
Region 27 uncorrelated					
bias	0.000	−0.001	−0.002	−0.002	0.002
RMSE	0.015	0.011	0.008	0.005	0.006
Region 27 correlated					
bias	0.000	−0.001	−0.002	−0.002	0.002
RMSE	0.025	0.019	0.015	0.008	0.010
Region 60 quantiles	0.610	0.703	0.778	0.991	1.174
Region 60 uncorrelated					
bias	−0.003	−0.021	−0.018	−0.001	−0.005
RMSE	0.027	0.031	0.025	0.006	0.011
Region 60 correlated					
bias	0.001	−0.016	−0.016	0.000	−0.006
RMSE	0.039	0.037	0.028	0.009	0.015
Region 82 quantiles	0.440	0.512	0.587	0.948	1.277
Region 82 uncorrelated					
bias	0.000	−0.001	−0.001	−0.001	0.001
RMSE	0.015	0.011	0.009	0.006	0.006
Region 82 correlated					
bias	−0.001	−0.002	−0.002	−0.002	0.001
RMSE	0.032	0.024	0.020	0.012	0.013

TABLE 5. Simulation results for January precipitation.

Probability	0.02	0.05	0.10	0.50	0.80
Region 10 quantiles	0.241	0.331	0.459	0.920	1.351
Region 10 uncorrelated					
bias	0.003	0.000	−0.002	0.002	−0.006
RMSE	0.008	0.007	0.008	0.006	0.009
Region 10 correlated					
bias	0.003	0.002	0.000	0.002	−0.007
RMSE	0.029	0.028	0.030	0.020	0.026
Region 106 quantiles	0.250	0.355	0.466	0.927	1.381
Region 106 uncorrelated					
bias	0.004	0.002	0.000	−0.002	0.000
RMSE	0.018	0.013	0.010	0.007	0.007
Region 106 correlated					
bias	0.003	0.001	0.000	−0.002	0.000
RMSE	0.044	0.034	0.028	0.017	0.019
Region 49 quantiles	0.067	0.164	0.281	0.845	1.516
Region 49 uncorrelated					
bias	0.004	0.003	0.002	0.000	−0.002
RMSE	0.019	0.014	0.011	0.012	0.008
Region 49 correlated					
bias	0.006	0.005	0.003	−0.002	−0.002
RMSE	0.042	0.032	0.028	0.025	0.019
Region 27 quantiles	0.030	0.076	0.156	0.728	1.631
Region 27 uncorrelated					
bias	0.000	0.002	0.003	0.002	−0.009
RMSE	0.009	0.009	0.012	0.018	0.024
Region 27 correlated					
bias	0.000	0.002	0.003	0.003	−0.010
RMSE	0.017	0.018	0.023	0.037	0.039
Region 60 quantiles	0.224	0.282	0.376	0.978	1.414
Region 60 uncorrelated					
bias	−0.003	−0.001	0.001	−0.005	−0.003
RMSE	0.025	0.021	0.021	0.019	0.029
Region 60 correlated					
bias	−0.004	0.002	0.005	−0.007	0.000
RMSE	0.037	0.030	0.031	0.028	0.045
Region 82 quantiles	0.060	0.145	0.255	0.817	1.537
Region 82 uncorrelated					
bias	0.003	0.003	0.003	0.000	−0.002
RMSE	0.022	0.016	0.013	0.016	0.009
Region 82 correlated					
bias	0.001	0.002	0.001	−0.002	−0.001
RMSE	0.045	0.034	0.029	0.031	0.020

judged from Monte Carlo simulation, is high, especially for the left tail of the distribution. A regional quantile value can be easily transformed into a precipitation amount simply by multiplying the value by the regional mean precipitation. Site precipitation can be similarly estimated by multiplying the site mean by the regional quantile value. The regional values can therefore be used to estimate site precipitation for sites that did not enter into the quantile value computations if a mean for the site is known and if the site is assumed to have a precipitation climate that is similar to that of the region. Note that mean precipitation is used as a scale factor; we recommend, however, that the median, and not the mean, be used as a measure of central tendency because precipitation distributions are often skewed.

Adopting the mixed probability model described above necessitated the computation of L-moments for both the nonzero data and the total data (including the zero precipitation). The L-moments of the nonzero data were used for the distribution fitting, while the total data were used for homogeneity assessment. In the interest of reducing computer processing time, software was written to compute one set of L-moments from the other. It was found that computing the set for the nonzero data from the set for the total data led to numerical instability. The problem was rectified, however, by reversing the procedure so that L-moments for the nonzero data were used to compute those for the total data.

Since the L-moment methodology for assessing homogeneity and distribution goodness of fit makes extensive use of Monte Carlo techniques, the question arose as to how many simulations are required to produce stable results. The answer to the question is important in the sense that decisions were made, based on test results, that ultimately affect the end product—quantile values.

Initially, 500 simulations were run for the Monte Carlo processes. The input L-moments for all regions, durations, and starting months were processed on an IBM RISC 6000 workstation using the built-in uniform random number generator. The input data were also processed on an IBM 3090 with the same software except for the random number generator. Comparison of the output from the two parallel runs indicated that about 3% of the decisions based on the values of the homogeneity and goodness-of-fit statistics differed. Convergence of the Monte Carlo procedures to a "true" result was therefore questioned. The number of simulations was increased to 1000 and the data for selected regions were reprocessed on both computers. Doubling the number of simulations reduced the decision differences to about 1%. Once again the number of simulations was doubled to 2000. All of the data were then reprocessed in parallel. Except for a few isolated cases, the results from both computers were identical. Although not the same in every case, 2000 simulations were accepted as a reasonable compromise between convergence and computer processing time.

The use of L-moments is both labor and computer intensive. The labor is devoted mainly to defining homogeneous regions. Since the procedures assume that data observed at all locations in a region are random representations of a physical process that is operating everywhere in the region, it is incumbent upon the analyst to define regionalization criteria that reflect the physical processes and applications for which the data are to be analyzed. The computer intensivity results from the extensive use of the Monte Carlo simulation for fitting distributions and for computing homogeneity, goodness of fit, and confidence measures.

Computation of quantile values for the *National Drought Atlas* is the first known large-scale application of L-moments. Bridging the gap between theory and application was not as smooth as desired. When trying to apply the theory, many questions arose concerning regionalization, homogeneity, distribution fitting, and confidence measures. Answering these questions led not only to easily used probability statements that re-

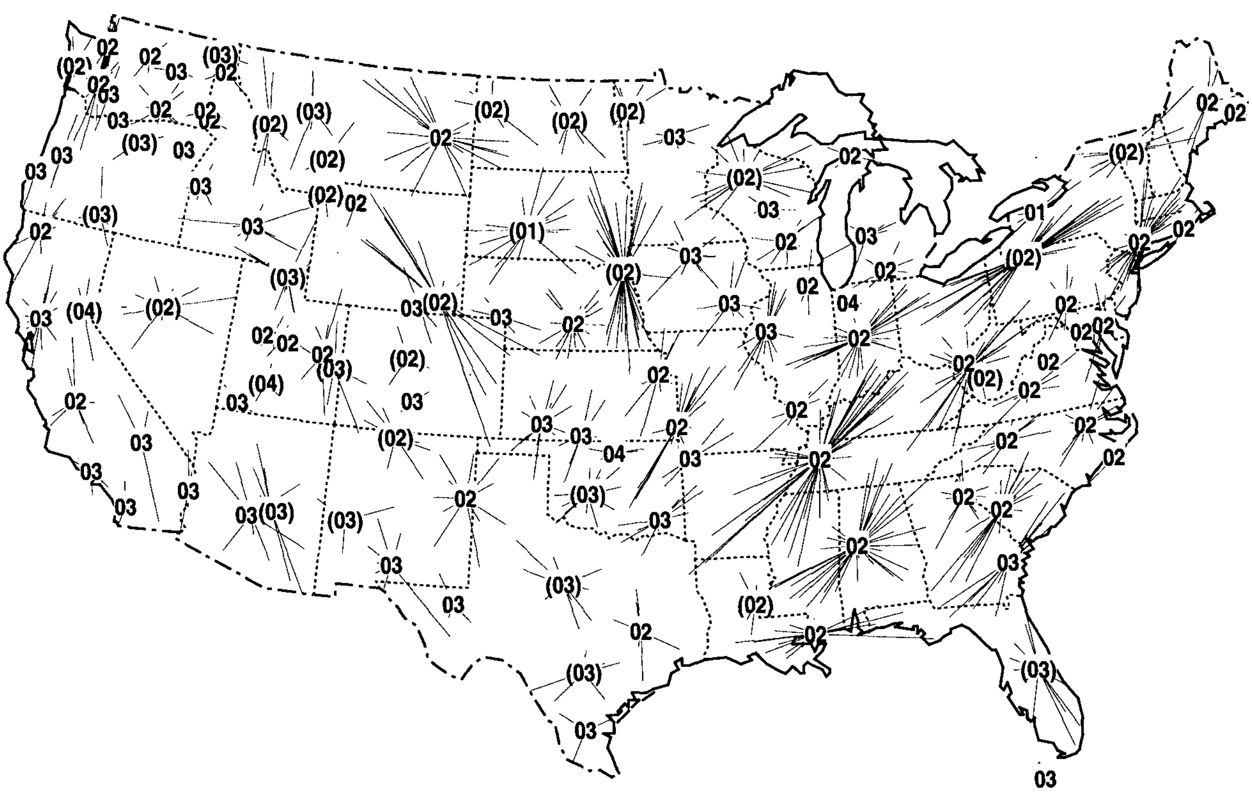

FIG. 16. Root-mean-square error (RMSE) of the ratio of the December through January three-month total precipitation 0.10 to the median quantile values taking intersite correlation into account. The number in the centroid of a region indicates the RMSE; lines radiating from the centroid end at the location of stations within a region. Heterogeneous regions are indicated by parentheses around the RMSE.

flect underlying physical processes, but also to refinement of the theory, modification of initial software, and development of new software.

REFERENCES

Guttman, N. B., 1989: Statistical descriptors of climate. *Bull. Amer. Meteor. Soc.*, **70**, 602–607.

——, 1993: The use of L-moments in the determination of regional precipitation climates. *J. Climate*, **6**, 2309–2325.

——, W. J. Werick, G. E. Willeke, and W. O. Thomas, Jr., 1991: A drought atlas of the United States. *Proc. of the Seventh Conf. on Applied Climatology,* Salt Lake City, UT, Amer. Meteor. Soc., 233–237.

Hosking, J. R. M., 1986: The Wakeby distribution. Research Report RC 12302, IBM Research Division, 21 pp. [Available from IBM Thomas J. Watson Research Center, Distribution Services F-11 Stormytown, P.O. Box 218, Yorktown Heights, NY 10598.]

——, 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc. B*, **52**, 105–124.

——, 1991: Fortran routines for use with the method of L-moments, Version 2. Research Report RC 17097, IBM Research Division. [Available from IBM Thomas J. Watson Research Center, Distribution Services F-11 Stormytown, P.O. Box 218, Yorktown Heights, NY 10598.]

——, and J. R. Wallis, 1988: The effect of inter-site dependence on regional flood frequency analysis. *Water Resour. Res.*, **24**, 588–590.

——, and ——, 1990: Regional flood frequency analysis using L-moments. Research Report RC 15658, IBM Research Division, 12 pp. [Available from IBM Thomas J. Watson Research Center, Distribution Services F-11 Stormytown, P.O. Box 218, Yorktown Heights, NY 10598.]

——, and ——, 1991: Some statistics useful in regional frequency analysis. Research Report RC 17096, IBM Research Division, 23 pp. [Available from IBM Thomas J. Watson Research Center, Distribution Services F-11 Stormytown, P.O. Box 218, Yorktown Heights, NY 10598.]

Kite, G. W., 1988: *Frequency and Risk Analyses in Hydrology.* Water Resources Publications, 257 pp.

Kotz, S., N. L. Johnson, and C. B. Read, Eds. 1988: "Wakeby distributions," in *Encyclopedia of Statistics, vol. 9,* Wiley, 513–514.

Landwehr, J. M., N. C. Matalas, and J. R. Wallis, 1979: Estimation of parameters and quantiles of Wakeby distributions. *Water Resour. Res.*, **15**, 1361–1379, 1672.

Wallis, J. R., 1989: Regional frequency studies using L-moments. *Research Report RC 14597,* IBM Research Division, Yorktown Heights, NY. [Available from IBM Thomas J. Watson Research Center, Distribution Services F-11 Stormytown, P.O. Box 218, Yorktown Heights, NY 10598.]

Willeke, G. E., N. B. Guttman, and W. O. Thomas, Jr., 1991: A national drought atlas for the United States of America. *Proc. of the United States–People's Republic of China Bilateral Symposium on Droughts and Arid-Region Hydrology,* Tucson, AZ. Open File Report 91-244. U.S. Geological Survey, 45–50.